Table of Contents

Computational Linguistics

Semantics

Invited Talk:

Multiword Expressions: A Pain in the Neck for NLP 1
Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger

A Hypothesis on the Origin of the Sign Types 16
Roland Hausser

Quantification and Intensionality in Situation Semantics 32
Roussanka Loukanova

Generalized Quantification in Situation Semantics 46
Roussanka Loukanova

Sign Language Translation via DRT and HPSG 58
Éva Sáfár and Ian Marshall

Multilayered Extended Semantic Networks as a Language for Meaning Representation in NLP Systems 69
H. Helbig and C. Gnörlich

Constructing a Sensuous Judgment System Based on Conceptual Processing 86
Atsushi Horiguchi, Seiji Tsuchiya, Kazuhide Kojima, Hirokazu Watabe, and Tsukasa Kawaoka

Towards a Natural Language Driven Automated Help Desk 96
Melanie Knapp and Jens Woch

Word Sense Disambiguation

Invited Talk:

Lexical Tuning 106
Yorick Wilks and Roberta Catizone

A Baseline Methodology for Word Sense Disambiguation 126
Ted Pedersen

An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet 136
Satanjeev Banerjee and Ted Pedersen

Feature Selection Analysis for Maximum Entropy-Based WSD 146
Armando Suárez and Manuel Palomar

Combining Supervised-Unsupervised Methods
for Word Sense Disambiguation 156
Andrés Montoyo, Armando Suárez, and Manuel Palomar

A Proposal for WSD Using Semantic Similarity 165
Susana Soler and Andrés Montoyo

Anaphora

Invited Talk:

A New, Fully Automatic Version
of Mitkov's Knowledge-Poor Pronoun Resolution Method 168
Ruslan Mitkov, Richard Evans, and Constantin Orasan

Pronominal Anaphora Generation in an English-Spanish MT Approach .. 187
Jesús Peral and Antonio Ferrández

Using LSA for Pronominal Anaphora Resolution 197
Beata Klebanov and Peter Wiemer-Hastings

Syntax and Parsing

The Spanish Auxiliary Verb System in HPSG 200
Ivan Meza and Luis Pineda

Surface Syntactic Relations in Spanish 210
Igor A. Bolshakov

Parsing Ill-Formed Inputs with Constraint Graphs 220
Philippe Blache and David-Olivier Azulay

Part of Speech Tagging

Part-of-Speech Tagging with Evolutionary Algorithms 230
Lourdes Araujo

Formal Methods of Tokenization for Part-of-Speech Tagging 240
Jorge Graña, Fco. Mario Barcala, and Jesús Vilares

Sepe: A POS Tagger for Spanish 250
Héctor Jiménez and Guillermo Morales

Fuzzy Set Tagging 260
Dariusz J. Kogut

Lexicon and Corpus

Invited Talk:

Towards a Standard for a Multilingual Lexical Entry:
The EAGLES/ISLE Initiative 264
Nicoletta Calzolari, Antonio Zampolli, and Alessandro Lenci

Quantitative Comparison of Homonymy in Spanish EuroWordNet
and Traditional Dictionaries 280
Igor A. Bolshakov, Sofia N. Galicia-Haro, and Alexander Gelbukh

Compilation of a Spanish Representative Corpus 285
Alexander Gelbukh, Grigori Sidorov, and Liliana Chanona-Hernández

Aligning Multiword Terms Using a Hybrid Approach 289
Arantza Casillas and Raquel Martínez

Automatic Selection of Defining Vocabulary
in an Explanatory Dictionary 300
Alexander Gelbukh and Grigori Sidorov

Text Generation

Integrated Natural Language Generation
with Schema–Tree Adjoining Grammars 304
Karin Harbusch and Jens Woch

Experiments with a Bilingual Document Generation Environment 314
Arantza Casillas and Raquel Martínez

Morphology

A Computational Model of Change in Politeness
with the Addition of Word Endings 318
Tamotsu Shirado and Hitoshi Isahara

Tartar Morphology Implementation 327
Dj.Sh. Suleymanov

Speech

Automatic Generation of Pronunciation Lexicons for Spanish 330
Esmeralda Uraga and Luis Pineda

Intelligent Text Processing

Spelling

Diacritics Restoration: Learning from Letters
versus Learning from Words 339
Rada F. Mihalcea

Information Extraction and Information Retrieval

A Comparative Study of Information Extraction Strategies 349
Ronen Feldman, Yonatan Aumann, Michal Finkelstein-Landau, Eyal Hurvitz, Yizhar Regev, and Ariel Yaroshevich

Answer Extraction in Technical Domains 360
Fabio Rinaldi, Michael Hess, Diego Mollá, Rolf Schwitter, James Dowdall, Gerold Schneider, and Rachel Fournier

Automatic Extraction of Non-standard Lexical Data for a Metalinguistic Information Database 370
Carlos Rodríguez

Text Segmentation for Efficient Information Retrieval 373
Fernando Llopis, Antonio Ferrández, and José Luis Vicedo

Using Syntactic Dependency-Pairs Conflation to Improve Retrieval Performance in Spanish 381
Jesús Vilares, Fco. Mario Barcala, and Miguel A. Alonso

Summarization

Multi-document Summarization Using Informative Words and Its Evaluation with a QA System 391
June-Jei Kuo, Hung-Chia Wung, Chuan-Jie Lin, and Hsin-Hsi Chen

Text Mining

Automated Selection of Interesting Medical Text Documents by the TEA Text Analyzer 402
Jan Žižka and Aleš Bourek

Text Classification and Categorization

Chinese Documents Classification Based on N-Grams 405
Shuigeng Zhou and Jihong Guan

Cross-Lingual Document Similarity Calculation Using the Multilingual Thesaurus EUROVOC 415
Ralf Steinberger, Bruno Pouliquen, and Johan Hagman

Empirical Formula for Testing Word Similarity and Its Application for Constructing a Word Frequency List 425
Pavel Makagonov and Mikhail Alexandrov

Document Processing

AutoMarkup: A Tool for Automatically Marking up Text Documents 433
Shazia Akhtar, Ronan G. Reilly, and John Dunnion

Identification of Recurrent Patterns to Extract Definitory Contexts 436
Gerardo Sierra and Rodrigo Alarcón

Demo Descriptions

Specification Marks Method: Design and Implementation 439
Sonia Vázquez, M^a Carmen Calle, Susana Soler, and Andrés Montoyo

Author Index 443